GUIDANCE FOR BENCHMARK DOSE (BMD) APPROACH - QUANTAL DATA

DPR MT-1

Health Assessment Section
Medical Toxicology Branch

Department of Pesticide Regulation
September 2004

ACKNOWLEDGEMENT

This Document is a product of the Health Assessment Section (HAS) in Medical Toxicology Branch, Department of Pesticide Regulation, Cal/EPA.

This document should be cited as:

> *DPR MT-1, 2004. Guidance for Benchmark Dose (BMD) Approach - Quantal data. Medical Toxicology Branch, Department of Pesticide Regulation, California Environmental Protection Agency, Sacramento, CA.*

The following staff (alphabetically listed) contributed to the production of this document: C. Jenks, Ph.D., E. Kwok, Ph.D., L. Lim, Ph.D., D.A.B.T., and N. R. Reed, Ph.D., D.A.B.T. (Lead person).

LIST OF ABBREVIATIONS

AIC	Akaike's Information Criterion
BMD	Benchmark Dose
BMR	Benchmark Dose Response
Cal/EPA	California Environmental Protection Agency
ChE	cholinesterase
DPR	Department of Pesticide Regulation
ED	Effective Dose at a specified response level (e.g., ED_{05}: Effective Dose at 5% response or BMR); also referred to as BMD (e.g., BMD_{05})
HAS	Health Assessment Section of Medical Toxicology Branch, DPR
LED	Lower bound of ED (e.g., LED_{05}: lower 95^{th} confidence bound of ED_{05}); also referred to as BMDL (e.g., $BMDL_{05}$)
LOEL	Lowest-Observed-Effect Level
MOE	margin of exposure
NOEL	No-Observed-Effect Level
OP	organophosphate
PB/PK model	physiologically based pharmacokinetic model
RPF	relative potency factor
USEPA	United State Environmental Protection Agency

I. INTRODUCTION

This document provides the necessary background and guidance for a consistent application of the benchmark dose (BMD) approach in the dose-response assessment of quantal data. It does not include guidance for the analysis of "nested data" that are most commonly seen in reproductive and developmental endpoints when the response of fetuses from one litter are inter-related. Crucial scientific issues in this document underwent a series of discussions and deliberation within the Health Assessment Section (HAS) of Medical Toxicology Branch to ensure sound scientific considerations. The guidance for continuous data analysis is available in a parallel document, *Guidance for Benchmark Dose (BMD) Approach - Continuous Data* (DPR MT-2, 2004).

The current risk assessment practice assumes that a threshold dose exists for effects other than oncogenicity, i.e., toxicologically significant effects are not likely to occur below the threshold dose. Two approaches can be used to define this threshold dose. The traditional approach determines the toxicity threshold as the no-observed-effect level (NOEL). NOEL is the highest dose in a study at which no effects are established (i.e., observed or measured). The next higher dose at which effects are seen is the lowest-observed-effect Level (LOEL). These NOELs and LOELs may be established based on statistically significant responses (e.g., $p \leq 0.05$) at the LOEL or by the evidence of a continuum of response with increasing dose. In this approach, the determination of the threshold dose is dictated by the dose selection in a toxicity study.

An alternative to the NOEL-LOEL approach is the BMD approach. It involves fitting a mathematical model to the entire dose-response dataset for an endpoint, and allowing the model to estimate the threshold dose corresponding to a level of benchmark response (BMR). This BMR is set at a certain level (e.g., 1%, 5%, 10%) as defined by the risk assessor. The BMD is either the model's best estimate of the effective dose (ED) at the BMR or the statistical 95th percent lower bound of ED (LED). Accordingly, BMD can be expressed as ED_{01}, ED_{05}, ED_{10}, or LED_{01}, LED_{05}, LED_{10}. Other comparable terms have also been used, such as BMD_{01}, BMD_{05}, BMD_{10}, and $BMDL_{01}$, $BMDL_{05}$, $BMDL_{10}$.

The NOEL-LOEL approach is relatively simple in that they can be determined directly from a study. Conversely, the BMD approach requires an extra step of fitting models to the dose-response data before determining the EDs and LEDs. However, the NOEL-LOEL approach has several limitations. It tends to focus only on data points at the apparent NOEL and LOEL, and not making full use of the entire dataset. This could result in different NOELs for an endpoint from two "identical" studies that differ only by the choice of dose for study. The NOEL-LOEL approach also tends to "reward" studies with smaller sample size or greater variations in endpoint measurement by assigning a higher NOEL based on statistical comparison to the controls. In reality, data from this type of study could mean greater uncertainty and higher probability for a false negative. Moreover, when a NOEL cannot be directly determined from a study (e.g., effects are present at the lowest tested dose), the NOEL-LOEL approach is inadequate to define a threshold for risk assessment. The current default practice is to divide the LOEL by a somewhat arbitrary uncertainty factor, usually within 10 (e.g., 1, 3, 10). Given the same dataset, the BMD approach can overcome these limitations of the NOEL-LOEL approach.

The advantages of the BMD approach are summarized below:

- Characterize the dose-response curve by using all pertinent data points
- Allow consistency in establishing the threshold dose from all studies and chemicals (i.e., corresponding to a given response level for an endpoint)
- Account for the greater uncertainty due to smaller sample size or greater variation in endpoint measurements or observations when the threshold dose is established as the LED
- Consistently estimate the threshold dose when no NOEL can be established (i.e., the lowest tested dose is the LOEL)

Although the BMD approach was introduced in the 90's, it has not been widely used until recently. One of the reasons is that suitable mathematical models for the variety of data types were limited and costly. This obstacle was removed recently with the BMD software made publicly available by USEPA in 2001 and subsequently updated. The other reason was the need for consistent criteria in the application of BMD approach. These include: the choice of model, the criteria for use of data in modeling, and the choice of BMR and BMD for risk assessment. The guidance for these areas is provided in this document.

II. GENERAL GUIDANCE

Prior to applying the benchmark dose (BMD) modeling, all toxicity endpoints should be identified in the Hazard Identification phase and accompanied with the identification of the no-observed-effect level (NOELs) (if possible) and the lowest-observed-effect level (LOEL).

To make the full use of a BMD approach and consistently account for the differential qualities between studies (i.e., sample size, measurement or observational variations), the HAS consensus is to identify the risk assessment threshold dose as the lower bound of the effective dose (LED). Nevertheless, a presentation of both the ED (best estimate of the effective dose) and LED values in the dose-response assessment is advisable as they provide more thorough information on modeling. For uniformity, the DPR convention for the threshold dose is expressed in "LED" (e.g., LED_{01}, LED_{05}, LED_{10}) and not BMDL (lower bound of BMD, e.g., $BMDL_{01}$, $BMDL_{05}$, $BMDL_{10}$).

III. SUMMARY GUIDANCE

The HAS guidance is summarized below. Detail discussions for the modeling process pertinent to quantal data are presented in Section IV.

1. **Endpoint Selection:** Considerations are given to the effects which are toxicological significant and/or adverse. Data for several pertinent endpoints should be modeled to ensure finding the lowest BMD from all datasets. Alternatively, BMD approach can be applied on a per need basis (e.g., when no NOEL can be established)

2. **Data Criteria:** The modeling requires data of individual test subjects or their summarized form (group dose, size, response incidence). A dataset should have at least two treated groups other than the control, with either a significant change in response with increasing dose (positive trend; $p \leq 0.05$) or a significant pair-wise increase in response ($p \leq 0.05$) in at

least one treated group. Datasets with near maximum response at the lowest tested dose is generally not a good candidate because of the extensive extrapolation (see below).

3. **Uncertainties in Extrapolation:** Extensive extrapolation below the experimental range of response should be avoided because it tends to introduce greater uncertainties.

4. **Data Conversion:** Continuous or categorical data can be converted to quantal data when necessary, although this usually results in some loss of information (DPR MT-2, 2004).

5. **Choice of Model and Options:** When more than one model can adequately describe a dataset, the model with the best fit should generally be used. The model fit criteria and considerations for model selection are given in Section IV, Step 5. If none of the available models can fit, or the model fit is poor in the region near the BMR (e.g., disparity between the "observed" and the "estimated"), the data point(s) high above the BMR may be excluded, while still retaining the minimum number of dose group needed for modeling (see Section IV, Step 2). Physiologically based pharmacokinetic (PB/PK) model can also improve the model fit by estimating the dose or concentrations of the parent chemical or its active metabolites at the target site(s). When no model can adequately describe the dose response relationship, it may be necessary to revert to the NOEL-LOEL approach.

6. **Define BMR:** For characterizing the risk, the HAS default BMR is 5% "extra risk". Modifications of the response level (i.e., down to 1%; up to ≥10%) can be made based on the biological significance of the endpoint and other toxicological considerations. The default should not be used when the BMR is specified in other HAS guidelines for a particular endpoint. For comparing the relative toxicity based on the BMD of a same endpoint, the preference is to use ED instead of LED, and can consider using the same model if the fit is valid. When comparing the relative sensitivity of multiple endpoints, the BMR (e.g., 1, 5, or 10%) for each endpoint should be based on its severity and other toxicological considerations, and not necessarily have to be at the same level across all endpoints.

IV. DISCUSSIONS

This section provides more extensive discussion on the guidance presented in the previous section.

Step 1: Endpoint Selection

In the NOEL-LOEL approach, the critical endpoint selected for a study is the biologically or toxicologically significant effect with the lowest NOEL. The effects with the lowest LED can similarly be selected when using the BMD approach. There is no guarantee, however, that the effect with the lowest NOEL or LOEL also has the lowest ED or LED because these parameters are inherently dependent on the shape of the dose-response curve. Therefore, several endpoints may need to be modeled to determine which has the lowest LED (may or may not be at the same BMR - see Step 7 discussion) and is therefore the critical endpoint. Indiscriminately modeling all endpoints to determine the lowest LED could be burdensome and unnecessary (USEPA, 1995). A more focused approach to ensure capturing the lowest LED is to model endpoints having LOELs up to approximately 5-fold the lowest LOEL. For example, effects having

LOELs up to 50 mg/kg/day could be modeled if 10 mg/kg/day was the lowest dose with an established effect for the study. Endpoints without biological or toxicological significance or showing no dose-response relationship can be excluded (USEPA, 2000). For identifying the lowest LOEL or LED from all pertinent inhalation studies, the exposure concentration-duration (e.g., ppm, hours/day) could be expressed in the dose term (e.g., mg/kg/day) to account for the duration variable and species-specific breathing rate. A given level of BMR (e.g., 5% change) for one endpoint may also have different biological significance than for other endpoints. Hence, choosing the lowest LED as the critical threshold for risk assessment based on the same level of response across all endpoints may not be valid. The modeler needs to exercise his/her own judgment to determine the BMR level for each endpoint, and take into account the continuum of adverse effects with increasing dose, especially when a sensitive endpoint or biological biomarker is modeled.

Alternatively, the BMD approach can be applied only on a per need basis. For example, when no NOEL can be established in a study (i.e., toxicological effects are observed at the lowest tested dose). In this case, without a BMD approach, the current default for estimating a NOEL would be to scale down from the LOEL using a somewhat arbitrary uncertainty factor of up to 10. The BMD approach can also be used in the inter-study or inter-species comparison for a specific endpoint.

Step 2: Data Criteria

Endpoints to be modeled must have a complete set of quantal data which includes dose levels, group sizes, and response incidences. Data set for BMD modeling should have at least two data points other than the control and with either a significant change in response with increasing dose (positive trend; $p \leq 0.05$) or a significant pair-wise increase in response ($p \leq 0.05$) in at least one treated group (Barnes *et al.*, 1995). In general, datasets with near maximum response at the lowest tested dose would not be a good candidate for modeling because it requires extensive extrapolation to the BMD (see discussion below).

Step 3: Uncertainties in Extrapolation

In general, the closer the response levels of the treatment groups are to the BMR, the less effect the choice of model will have on the estimated ED and LED (USEPA, 1995). Extrapolation beyond the experimentally observable or measurable range is not recommended because different models can yield widely different LED (i.e., greater than a factor of 3) (Crump, 1984; USEPA, 2000). Thus, datasets with near maximum response at the lowest tested dose is generally not a good candidate because it requires extensive extrapolation.

When some degree of extrapolation is necessary (e.g. no NOEL can be determined from a study), consideration is given to the biological plausibility of the shape of the dose-response, especially in the low dose region.

Step 4: Data Conversion

Quantal data is dichotomous and consists of two categories: “affected” and “non-affected” (USEPA, 1995). However, observational data may be graded and consist of more than two

categories (e.g., "mild," "slight," "medium" or "severe"). For applying the quantal BMD approach, such data are converted into a quantal form by assigning responses to "affected" and "non-affected" categories. For example, in the risk assessment of antimony trioxide inhalation by the USEPA (IRIS, 1997), the incidence of chronic interstitial inflammation was reported in several categories, such as "non-affected," "minimal," "slight," and more severe designations. These categorical data were converted to quantal data by considering all affected categories as "affected." However, a high background incidence of "minimal" and "slight" grades of inflammation obscured the dose-response relationship of the more severe treatment related effect. The dose-response relationship was reclassified by placing only incidences of inflammation that were above "slight" into the "affected" category and grouping the remaining categories as "non-affected". Thus, the BMD was established based on more severe interstitial inflammation. In this case, a discussion on the criterion for toxicity threshold should accompany the analysis, and the severity factor be accounted for in the BMR determination.

Although the software for several continuous models are available (USEPA, 2003), there may be occasions when it is advantageous to converting continuous data to quantal data before a BMD analysis. In some cases, converting continuous data into quantal data may more directly address a specific definition of adverse response. For example, body weight measurements for individual animals could be converted to the incidence of animals that have more than 10% lowering of body weight, if this is the criterion for a significant response. Rendering data quantal may also facilitate comparisons among datasets if the majorities are in quantal form. The obvious disadvantage of continuous to quantal conversion is the loss of information about the magnitude of response.

Step 5: Choice of Model and Options

Mathematical Models: The available mathematical models are listed in Table 1. The models distributed by USEPA (http://cfpub.epa.gov/ncea/cfm/nceahome.cfm) are footnoted.

Model Run Options: The following options are available for the USEPA BMD Models:

- BMR Type: **"**Additional Risk" vs. "Extra Risk"
  "Additional risk" is the probability of response to the treatment beyond the background occurrence. "Extra risk" is the probability of response to the treatment among the animals that would not otherwise have shown the endpoint of interest (USEPA, 1995). For example, if treatment increases a background response from 90% to 91%, the "additional risk" would be 1% but the "extra risk" is 10%. If the background response is zero, both types of risk would have the same numerical value. "Extra risk" is the HAS default for the BMD approach.

$$\text{Addtional risk} = P(d) - P(0)$$

$$\text{Extra risk} = \frac{P(d) - P(0)}{1 - P(0)}$$

$P(d)$ = Response at dose d
$P(0)$ = Response of control

- Options for Restriction:
  - Curve smoothing options are "unique" (HAS default) or "csplines"
  - Beta coefficient for Multistage model can be set to ≥ 1
  - Power for Gamma and Weibull models can be set to ≥ 1
- Degree of polynomial (n): can be specified for Multistage model; the model default of "n = dose group minus 1" is the HAS default
- Log transformation of dose: Selecting this option allows the run of Log-Logistic and Log-Probit models

Table 1. Available Quantal Models

[1]Linear mean: $P(d) = b + s(d - d_0)$	[1]Power mean: $P(d) = b + s(d - d_0)^p$
[1]Logistic: $P(d) = \frac{1}{1 + e^{-(i+s(d-d_0))}}$	[2]Log-logistic: $P(d) = \frac{1-b}{1 + e^{-(i+s\ln(d-d_0))}}$
[2]Probit: $\text{Prob}\{d\} = N(i + s(d - d_0))$	[2]Log-probit (i.e., log-normal): $P(d) = b + (1-b)N(i + s\ln(d - d_0))$
[2]Quantal linear: $P(d) = b + (1-b)(1 - e^{-s(d-d_0)})$	[2]Quantal quadratic: $P(d) = b + (1-b)(1 - e^{-s(d-d_0)^2})$
[2]Weibull: $P(d) = b + (1-b)(1 - e^{-s(d-d_0)^p})$	[2]Multistage (i.e., quantal polynomial): $P(d) = b + (1-b)(1 - e^{-\sum_{j=1}^{n}\beta_j(d-d_0)^j})$
[2]Gamma Multi-Hit: $P(d) = b + (1-b)\frac{1}{G(\alpha)}\int_0^{s(d-d_0)} t^{p-1}e^{-t}dt$	

Symbols:

Probability of a response at dose d: $P(d)$	Intercept: i	Power: $p \geq 0$
Normal distribution function: N	Background: $0 \leq b < 1$	Slope: $s \geq 0$
Gamma function: G	Model parameters: β_n	Threshold: d_0

1/ Gephart *et al.*, 2001

2/ USEPA, 2003 – BMD software freely distributed by the USEPA. The 1.3.2 version does not allow specification of the presence of a threshold in the models.

Model Fit Criteria: The following criteria should be met for considering a model as adequate to describe the data.

- Model goodness-of-fit - χ^2 p value >0.05 (See: model output in Appendix A).
- Visual examination - inspect the graphical display for the model fit, especially when the goodness of fit p value is not available (Note: p value cannot be calculated when the degrees of freedom is <1).
- The χ^2 residual values - should not exceed $|2|$ (i.e., absolute value of 2) for each dose group, especially near the BMR.

Selection of Model: The best model is selected based on its accuracy in describing the data. Hence, it is recommended that all available models should be run and the model with the best fit would be used. An obvious consideration with the polynomial model is that, while the model fit may appear improved with increasing degree of the polynomial, the dose-response curve becomes "wavy" and lacks scientific support.

Considerations for selecting the final model for the BMD analysis are listed below.

- Consider the biological plausibility of the shape of the dose-response curve. Avoid over-parameterizing the model (see "Model Run Options" above).
- Consistency between the model estimated and the observed variables (See: model output in Appendix A), especially at the BMR and BMD region.
- Use the model estimates with care when the difference between ED and LED is great (i.e., the ED/LED ratio is large, e.g., >5).
- In general, the model that has the lowest AIC[1] (Akaike's Information Criterion) can be the model of choice (USEPA, 2003; Akaike, 1973; Stone, 1998). However, the selection of a final BMD model among multiple models with adequate fit should not rely solely on the AIC in its current form (Sand *et al*., 2002). Instead, all of the above considerations should be taken in determining the final model for the BMD analysis.
- In addition to the above criteria, considerations may be given for using a same model for similar datasets, e.g., data from males and females, or for comparing the potency of a groups of chemicals with the same mode of action and based on the same endpoint (see: Step 6 below).
- Comparing output from more than one valid models and simple sensitivity analysis (e.g., varying sample size or response rate of key data points) are also helpful for a better understanding of the modeling behavior for a particular set of data. In general, a small sample size widens the model confidence bounds especially toward lower BMR range (e.g. 1%). Murrell *et al* (1998) reported that the depending on the model and dataset, LED could be lower than the NOEL by 2-3 orders of magnitude.

Improve Model Fit: When the dose-response relationship shows stages of changing slope (e.g., plateau at the high dose range), a better fit at the range near BMR may be achieved by using a more flexible model such as the multistage model with increasing mathematical complexity (i.e., degree of polynomial). However, a simpler model is generally preferred over a more complex one with comparable fit, as model simplicity is

[1] The AIC takes into account both the model fit and the model complexity.

included in the AIC calculation (USEPA, 2000). When a full dataset cannot be adequately described with a model (especially if there is a plateau or decreased response with increasing dose), it may be reasonable to focus on the lower dose region, when it is more relevant to the BMR. This can be done by excluding the highest dose data point(s) far above the BMR, while still maintain at least three data points (i.e., controls plus two treatment groups) with a positive trend (i.e., increase in response).

Physiologically based pharmacokinetic (PB/PK) model can also be used to refine the dose estimation for a better model fit. Estimating the dose or concentrations of the parent chemical or its active metabolites at the target site(s) through PB/PK model is especially useful when the plateau of dose-response relationship at the high dose range is due to saturation of metabolic processes or transport systems.

In the case when the response level drops with increasing dose, it is also important to verify that all affected animals are accounted for. For example, with increasing severity at higher doses, the response may have a different designation or classifications (e.g., "mild atrophy" becomes "severe atrophy"), giving the appearance of reduced response at the milder category. Thus, all test subjects should be accounted for in the overall incidence (see: discussions under Step 4). When no model can adequately describe the dose response relationship, it may be necessary to revert to the NOEL-LOEL approach.

Step 6: Define BMR

The final step is defining the BMR for which the BMD will be established. The HAS default for defining the BMR is based on "extra risk". Other areas of considerations for the choice of BMR are presented for the following three common applications of the BMD approach.

A. Threshold for Risk Assessment

Although one advantage of the BMD approach over the traditional NOEL-LOEL approach is the consistency in establishing the BMD across all studies and among chemicals, this does not necessarily mean that a single level of BMR should be used for all quantal endpoints. In fact, it is reasonable to allow flexibility in the BMR level based on the endpoints of concern, such that lower BMRs could be used for more severe or detrimental endpoints for the protection of human health.

Background

Several reports approached the issue of defining BMR from the standpoint of a comparison between the NOEL and the BMD (see: Gaylor, 1989, 1992; Chen and Kodell, 1989; Kimmel , 1990); Allen *et al*., 1994; Auton, 1994; Haag-Gronlund *et al*., 1995; Murrell *et al*., 1998; Fowles *et al*., 1999). When expressed as the NOEL-to-LED ratio, the closer the LED is to the NOEL, the nearer the NOEL/LED ratio is to one. Results from two reports (Allen *et al*., 1994; Fowles *et al*., 1999) are summarized in Table 2 for illustrating the varying NOEL/LED ratios due to modeling considerations (e.g., data format, choice of model) and for different endpoints.

A close proximity of NOELs to the BMD at 5-10% BMR has been shown in many reports. The example given in Table 2 from the analysis by Allen *et al* (1994) includes two types of data format (quantal versus continuous). The first set is quantal, based on the proportion of affected litter (i.e., number of litters with at least one affected fetus). The second set is continuous, based on the number of affected fetus per each litter. The higher NOEL/LED ratio for the first set could be due to both the quantal data format and quantal modeling. Quantalizing data gives equal weight to litters with any number of affected fetuses, resulting in reduced sensitivity for detecting dose-related response and possibly higher NOELs. Compared to continuous model, quantal models also tend to give lower estimates of EDs and LEDs (Allen *et al*., 1994). Similar comparison of the NOEL to BMD by Haag-Gronlund *et al* (1995) also showed NOELs closest to the LED_{10}.

Table 2. Comparison of the NOEL and BMD

Endpoints	$NOEL/LED_{01}$	$NOEL/LED_{05}$	$NOEL/LED_{10}$	References
Developmental (affected/total litter – quantal data)	29±44 (median: 19)	5.9±8.4 (median: 4)	2.9±3.9 (median: 2.0)	Allen *et al*., 1994[a]
Developmental (% affected fetus/litter – continuous data)	4.3±4.5 (median: 2.5)	1.2±0.88 (median: 0.96)	0.72±0.44 (median: 0.62)	Allen *et al*., 1994[b]
Acute toxicity	1.6±0.84	1.16±0.38	0.99±0.27	Fowles *et al*., 1999[c]
Acute toxicity	3.59±3.75	1.59±0.87	1.17±0.46	Fowles *et al*., 1999[d]

a/ NOELs based either on expert judgment or iterative trend test (removing the highest data point until no significant trend was present). Endpoints included fetal death and gross, visceral, and skeletal malformations in mice, rabbits, rats, or hamsters. Quantal Weibull model was used for the quantal data.

b/ NOELs was determined as above. Continuous power model was used for the continuous data.

c/ Based on acute lethality toxicity for 100 chemicals, using probit model. The corresponding LOEL/LED ratios were 2.74, 1.84, 1.52 at 1, 5 and 10% BMR.

d/ Based on acute lethality toxicity for 100 chemicals, using Weibull model. The corresponding LOEL/LED ratios were 7.78, 2.81, 1.91 at 1, 5 and 10% BMR.

Other NOEL/LED comparisons can be used to support a lower BMR (e.g. ~1%). For example, using a "model free" approach (i.e., a point estimate), Gaylor (1989, 1992) indicated that the treatment-related response at the developmental NOELs from 120 studies ranged up to 4.5%, with only about one-fourth of the cases exceeding a 1% response. The analysis by Chen and Kodell (1989) also showed that the LED_{01} was comparable to the NOELs for developmental endpoints (i.e., resorption, death, malformation). Therefore, the BMR of 1% was suggested for teratological effects. Data analyses by Fowles *et al* (1999) also support a range of BMR between

1 - 10% for a lethality endpoint (Table 2). Two quantal models were used in this illustration. The less variation in the NOEL/LED ratio between the BMR of 1% to 10% in their first set of analysis (third row of data in Table 2) was attributed to the steep dose-response relationship for lethality and the difference between probit and Weibull models (Fowles *et al.,* 1999).

While the aforementioned NOEL-to-LED comparisons have often been used as justification for BMR at 1-10%, USEPA recommended the consideration of a "point of departure" (POD) that is below 10% BMR (Barnes *et al*., 1995, USEPA, 2000) because it is a common range of detection for dichotomous data.

HAS guidance

Conceptually, the BMR (1, 5, or 10% etc.) should represent a response level of no significant concern. Accordingly, the corresponding BMD would be comparable to the NOEL, not the LOEL. Thus, other than the interspecies and inter-individual uncertainty factors commonly applied to the NOEL for estimating the reference dose (RfD) or setting an acceptable margin of safety (MOE), no additional uncertainty factors (e.g., for LOEL-to-NOEL extrapolation) will be needed (Kimmel, 1990; Gaylor, 1992).

It should be noted that the NOELs in the NOEL-to-LED comparison are dependent on the dose selection of a study, while the estimation of LED is often model-dependent. Therefore, although useful, the comparisons contain the uncertainties inherent in the NOEL and LED determinations. As such, how close LED is to the NOEL should not be viewed as definitive for defining an appropriate BMR (i.e., 1, 5, or 10% or above). On the other hand, neither should statistical power be the only factor for defining the BMR. Many toxicological studies are not statistically designed for detecting significant quantal responses at sufficiently low level deemed as health protective. For example, although one positive response out of 4 dogs or 10 rats may not be statistically higher than a zero incidence in the controls, these responses should not be dismissed if they are treatment-related and especially if the endpoints are severe (e.g., death).

Thus, besides the comparison to the NOEL and the statistical considerations, severity and/or adversity of the toxicity endpoint should be a factor in defining a BMR on a flexible scale, thereby allowing lower BMR for more detrimental effects. The flexible scale is important because the dichotomously categorized effects can vary in severity (“mild”, “moderate”, "severe”) (USEPA, 1994). Moreover, because of the difference in an organ's functional reserve (i.e., the capability to handle toxic assault), a level of response (e.g., hypertrophy) in one organ (e.g., liver) may not be as detrimental as it is in another organ (e.g., brain). Applying a fixed BMR to both cases may have significant consequences in terms of protecting public health (Bogdanffy *et al*., 2001).

"Severity" and "adversity" are not necessarily two distinct and unrelated criteria. Oftentimes, as the gradation of a response moves up toward the higher end of the severity scale, the distinction between "severity" and "adversity" tends to diminish. For example, it is unlikely that a “severe” effect is not “adverse” in itself or not associated with clearly adverse effects. On the other hand, it may be hard to objectively define "adversity" when the response is on the mild end of the severity scale. Historically, within the conventional NOEL-LOEL approach, the term NOAEL (no-observe-adverse-effect level) could be used to distinguish the "adverse" from the "non-

adverse" endpoint. However, the terminology only conveys the judgement of "adversity", not necessarily defines it. The often-cited definition of an "adverse effect" is "*a functional impairment(s) or a pathological lesions(s) which may affect the performance of the organism or which reduce its ability to respond to additional change.*" (Dourson and Stara, 1983). In the case of mild effects, the broadness of this "adversity" concept still leaves room for subjective judgement on "adversity". For example, effect that are subtle or precursory may arguably be defined as merely an indicator of exposure (e.g., a natural defense or compensatory response to the exposure) than the evidence of toxicity in itself (USEPA, 1994).

Since the "adversity' and "severity" scales are not always independent of each other, HAS has chosen to define the BMR primarily on the severity scale, and allowing flexibility for adversity considerations. Specifically, 5% BMR (thus, the BMD at LED_{05}) is the default for moderate effects. A lower BMR (i.e., 1%) can be justified for severe effects and a higher BMR (i.e., above 5%) can be justified for mild effects. It is possible that a review of other auxiliary toxicity data could move an effect from "moderate" category by itself into a "severe" category, and a downward adjustment of the BMR from 5% to 1% may then be applied.

The severity classifications proposed by DeRosa *et al*. (1985) and USEPA (1994) can be used as starting points for assigning the BMR. They are reproduced in Table 3, with minor modifications. In this table, biological responses are grouped into mild, moderate, and severe categories, and the effects listed within each category are arranged from the least to the most severe. As described previously, organs with functional reserve such as the lung, liver, and kidney may be more resilient to the chemical insult than those that can only compensate damages to a limited degree, such as the central nervous system. Hence, the severity designation presented in Table 3 should be viewed in the context of organs/systems, and the default BMR adjusted accordingly.

In addition to the above considerations, the application of the BMD requires a careful examination of the dose-response relationship and the modeling behavior to ensure a sound selection of the BMR for each endpoint and dataset. The aforementioned default scale of BMR should not be used rigidly (see: further discussions in Section V - Model Output).

In conclusion, BMR should not be chosen solely based on the statistical power inherent in a study but also consider factors such as severity of the endpoint and limitations and strength in the modeling. A comparison between the NOEL and LED could provide valuable perspectives but should not dictate the eventual choice of BMR for characterizing the dose-response. The BMR should represent a "no-effect" level. The current HAS default is a BMR of 5% for effects of moderate severity, with flexibility for scaling downward to 1% or upward to ≥10% based on the severity of effects and other auxiliary toxicity information pertinent to the "adversity" considerations. The LED corresponding to the defined BMR would be the toxicity threshold dose (i.e., "NOEL-equivalent") for the subsequent MOE and RfD calculations in risk characterization. The general default BMR presented in this document should not be used when the BMR is specified in other HAS guidelines for a particular endpoint (e.g. cholinesterase inhibition, or ChE inhibition).

Table 3. Consideration of Endpoint Severity for Benchmark Dose Approach

Severity Category	Biological/Toxicological Effects
Mild	
	* Changes in enzyme level (e.g., increase SGPT, SGOT activity) or other biochemical parameters (e.g., increased serum cholesterol) consistent with the possible mechanism of action (e.g., liver damage) without apparent pathological, clinical, or absolute organ weight changes
	* Proliferation or other changes in organelles (e.g., centrilobular hepatocyte vacuolization) consistent with the possible mechanism of action (e.g., centrilobular hepatocyte necrosis, liver damage) without other apparent effects
	* Hyperplasia (e.g., urinary bladder), hypertrophy (e.g., parotid salivary gland parenchyma), or atrophy (e.g., degenerative change in adrenal medulla) without significant change in organ weights[a]
	* Effects whose significance to the organism are not entirely known (e.g., urine stain)
Moderate	
	* Hyperplasia, hypertrophy, or atrophy with significant changes in absolute organ weights[a]
	* Cellular changes including cloudy swelling, hydropic change (e.g., liver), or fatty infiltration. (e.g., liver, kidney)
	* Neuropathy without apparent behavioral, sensory, or physiological changes (e.g., decrease dopamine contents in brain areas).
	* Degenerative or necrotic tissue changes without apparent decrement in organ function (e.g., mild degeneration and necrosis of hepatocytes)
Severe	
	* Neuropathy with a measurable change in behavioral activity, sensory ability, or physiological function (e.g., impaired avoidance reaction and retention of a learned task)
	* Necrosis, atrophy, hyperplasia, or hypertrophy with a detectable decrement of organ functions (e.g., renal tubular atrophy and necrosis, dilation of collecting tubules, necrosis of the edematous papilla, nephritis)
	* Evidence of fetotoxicity (e.g., increased incidence of runts)
	* Pathological changes with definite organ dysfunction (e.g., congested or hemorrhagic lungs)
	* Decreased reproductive capacity or reproductive dysfunction (e.g., decreased conception, altered estrous cycle)
	* Neuropathy with change in motor control, sensory ability, or behavioral function; loss of motor control sensory ability or behavioral functions (e.g., extreme debilitation weakness and lethargy)
	* Pronounced pathologic changes with severe organ dysfunction and/or long-term sequelae (e.g., chronic nephropathy).
	* Teratogenic effect (malformation) with or without accompanying maternal toxicity (e.g., focal liver necrosis in developing pups, severe vacuolization).
	* Death or significant shortening of lifespan

a/ It should be noted that, in some organs, hypertrophy may be the result of increased metabolism (e.g., liver). Likewise, organ atrophy (e.g., thymus) could be a secondary effect due to the nutritional status of the tested animals.

B. Relative Toxicity

Background

The ratio of two BMDs at a given BMR of an endpoint is often used as a measure of relative toxicity in expressing the gender and species sensitivity to a chemical. The concept of relative toxicity is also applicable for assessing the risk of exposure to a mixture of chemicals with the same mode of action. For example, in assessing the risk of exposure to multiple organophosphate (OP) pesticides, USEPA estimated the Relative Potency Factors (RPFs) of more than 20 OPs based on the ratio of their BMDs to the BMD of an index OP chemical (in this case, methaminophos) at 10% brain ChE inhibition in female rats (USEPA, 2002). These RPFs were then used to scale and sum the exposure from all OPs for calculating the overall margin of exposure (MOE)[2].

HAS guidance

When comparing the sensitivity of a specific endpoint among studies, species, or chemicals at a specified BMR, it is desirable to base the comparison on the ED rather than the LED. The use of the ED avoids the uncertainty associated with the model-dependent tendency in the LED estimation, especially when more than one model is used to estimate the BMDs across all chemicals in the comparison. To avoid other model-dependent issues (e.g., the model-specific shape of the dose-response curve), it may also be useful to consider applying one common model to the same endpoint (i.e., with presumed same mode of action) for all chemicals in the comparison if the model fit is meets the criteria of good fit (see: Step 5 above).

C. Endpoint Sensitivity Comparison

Background

In risk assessment, the most sensitive endpoint is often defined as the endpoint that has the lowest threshold. In the NOEL-LOEL approach, this is usually the endpoint with the lowest NOEL. For the BMD approach, this is the endpoint with the lowest LED at a pre-determined BMR.

HAS Guidance

When comparing the relative sensitivity of multiple endpoints, the BMR (e.g., 1, 5, or 10%) for each endpoint should be based on its severity and other toxicological considerations, and not necessarily have to be at the same level across all endpoints.

V. MODEL OUTPUT

A sample text and graphic output for a 10% "relative deviation" by the Hill model is given in Appendix A. It provides information on the parameter estimates, the statistical tests, and the "BMD" (e.g., ED_{10}) and "BMDL" (e.g., LED_{10}). The Help manual (USEPA, 2003) should be consulted for further information.

[2] Margin of Exposure is the ratio of the toxicity threshold (e.g., NOEL, BMD) to the exposure.

For a better understanding of the impact of BMD modeling on the determination of toxicological thresholds, examples are presented in Table 4 for a diverse application of the BMD methodology. The salient illustrative points are highlighted below. These comments serve only as points of consideration, and should in no way direct or limit the wider exercise necessary for understanding the modeling approaches. On a case by case basis, some of these considerations may be included in risk assessment document to inform the support for the final conclusion on the BMD.

Table 4. Examples of BMD Application[a]

Dataset/Endpoint[a]	Dose in mg/kg/day (incidence in parenthesis)							
	NOEL	LOEL	BMR=1%		BMR=5%		BMR=10%	
			ED	LED	ED	LED	ED	LED
1. Red Stained Urine	32 (0/6)	100 (4/6)	75	10	82	20	85	28
2. Abortions[b]	20 (0/16)	80 (5/16)	50	1.6	60	8.2	66	10
3. Anorexia	20 (3/18)	80 (10/18)	6.7	4.5	13	8.8	19	13
4. Pituitary hypertrophy	74 (4/20)	300 (14/19)	2.6	1.7	13	8.9	27	18
5. Lethargy	<500 (0/30)	500 (1/30)	280	170	520	340	710	480
6. Ovary hyperplasia	11 (22/60)	55 (38/60)	7.2	5.8	16	13	23	19
7. Enlarged liver	15 (2/63)	75 (17/63)	19	12	34	23	47	33

a/ Data from an anonymous pesticide. Each set of BMD represents output from the model with the best fit.
1. Log-Logistic model - Pregnant rabbits received gavage exposure for 13 days
2. Gamma model - Pregnant rabbits received gavage exposure for 13 days
3. Log-Probit model - Pregnant rabbits received gavage exposure for 13 days
4. Quantal-Linear model - Rats received dietary exposure for 3 weeks
5. Long-Probit model - Male mice received a single gavage exposure
6. Quantal-Quadratic model - Female rats received dietary exposure for 24 months
7. Long-Probit model - Male mice received dietary exposure for 24 months

b/ The LED_{10} is 3-fold higher if the sample size is increased by 2-fold. See text for other considerations

- **Disparity between LED and NOEL-LOEL**: The significant differences between the LED and the NOEL or LOEL for datasets #1 through #5 appear to warrant a closer look at the modeling.

- **Sample size**: The small sample size for datasets #1 and #2 contributed to the wider 95th confidence bound from the best estimates. For example, the LED_{10} of dataset #2 increases 3-fold (to 30 mg/kg/day) if the sample size is proportionally doubled (from 16 to 32). Caution should be taken in interpreting the results, especially if the small sample size is typical for the type of study and when making comparison between studies with very different sample size.

  Some possible considerations are:
  - How does the sample size affect the LEDs and the model fit for different models?
  - Given that the plot showing great deviation between ED and LED curves for the gamma model (the chosen model for its lowest AIC) for dataset #2, should the choice of model be based on AIC alone?
  - For this particular dataset, the estimated power (Table 1, parameter p) for the gamma model was extremely high (i.e., 18). Is it biologically plausible?

- **LED to NOEL ratio**: Datasets #3 and #4 showed much lower LEDs than the NOEL, especially at lower BMRs. It is noted that the background incidence (at the control group) is zero for dataset #4, however, the toxicology review established the NOEL at 74 mg/kg/day, with an incidence of 4/20 (i.e., 20% response). Similarly, the NOEL of 20 mg/kg/day for dataset #3 is associated with an incidence of 3/18 (i.e., 17% response), while the background incidence is only 1/20 (i.e., 5%) and that the increase in response is dose-related. In these cases, there may be reasons to more closely examine the rationale behind the NOEL selection.

- **No NOEL can be established**: Dataset #5 is an ideal condition for applying BMD approach when no NOEL can be established from the study (i.e., significant response is noted at the lowest dose tested) and that the response at the LOEL of 500 mg/kg/day is relatively low (3%). It is likely that using the traditional default to estimate the NOEL by dividing the LOEL by 10 would have been unnecessarily health conservative.

VI. REFERENCES

Allen, B., R. J. Kavlock, C. A. Kimmel, E. M. Faustman. 1994. Dose-response assessment for developmental toxicity II. Comparison of generic BMD estimates with no observed adverse effect levels. *Fund Appl Toxicol* 23: 487-495.

Akaike, H. 1973. "Information theory and an extension of the maximum likelihood principle." In: Proceedings of the Second International Symposium on Information Theory, B. N. Petrov and F. Csaki, Eds. Akademiai Kiado, Budapest, pp. 267-281.

Auton, T. R. 1994. Calculation of benchmark dose from teratology data. *Regul Toxicol Pharmacol*. 19:152-167.

Barnes, D. G., G. P. Daston, J. S. Evans, A. M. Jarabek, R. J. Kavlock, C. A. Kimmel, C. Park, and H. L. Spitzer. 1995. Benchmark Dose Workshop: Criteria for use of a benchmark dose to estimate a reference dose. *Regul Toxicol Pharmacol* 21:296-306.

Bogdanffy, M., G. Daston, E. M. Faustman, C. A. Kimmel, G. L. Kimmel, J. Seed, and V. Vu. 2001. Harmonization of cancer and noncancer risk assessment: Proceedings of a consensus-building workshop. *Toxicol Sci* 61:18-31.

Chen , J. J. and R. H. Kodell. 1989. Quantitative risk assessment for teratological effects. *J Amer Stat Assoc* 84:966-971.

Dourson, M. L. and J. F. Stara. 1983. Regulatory history and experimental support of uncertainty (safety) factors. *Regul Toxicol Pharmacol* 3:224-238.

Fowles, J. R., G. V. Alexeeff, and D. Dodge. 1999. The use of benchmark dose methodology with acute inhalation lethality data. *Regul Toxicol Pharmacol* 29:262-278.

Gaylor, D. W. 1989. Quantitative risk analysis for quantal reproductive and developmental effects. *Environ Health Persp* 79:243-246.

Gaylor, D. W. 1992. Incidence of developmental defects at the no observed adverse effect level (NOAEL). *Regul Toxicol Pharmcol* 15:151-160.

Gephart, L. A., W. F. Salminen, M. J. Nicolich, and M. Pelekis. 2001. Evaluation of subchronic toxicity data using the benchmark dose approach. *Regul Toxicol Pharmacol* 33:37-59.

Haag-Gronlund, M., R. Fransson-Steen, and K. Victorin. 1995. Application of the benchmark method to risk assessment of trichloroethene. *Regul Toxicol Pharmacol* 21:261-269.

IRIS (Integrated Risk Information System). 1997. U.S. Environmental Protection Agency, . Cincinnati, OH.

Kimmel, C. A. 1990. Quantitative approaches to human risk assessment for noncancer health effects. *Neurotoxicology* 11:189-198.

Murrell, J., C. J. Portier, and R. W. Morris. 1998. Characterizing dose-response I : Critical assessment of the benchmark dose concept. *Risk Anal* 18:13-26.

Stone, M. 1998. Akaike's Criteria. In: Encyclopedia of Biostatistics. Volume 1, A-COX. Armitage, P. and T. Colton, eds. John Wiley and Sons, N.Y.

USEPA. 1994. Methods for derivation of inhalation reference concentrations and application of inhalation dosimetry. Environmental Criteria and Assessment Office. Office of Health and Environmental Assessment. Office of Research and Development. U.S. Environmental Protection Agency, Washington DC. EPA/600/8-90/066F.

USEPA. 1995. The use of the benchmark dose approach in health risk assessment. Risk Assessment Forum, Office of Research and Development, U.S. Environmental Protection Agency, Washington DC.

USEPA. 2000. Benchmark dose technical guidance document. external review draft. Rick Assessment Forum, U.S. Environmental Protection Agency,Washington, DC. EPA/630/R-00/001.

USEPA. 2003. Help manual for benchmark dose software Version 1.3.2. Office of Research and Development, Washington DC.

Appendix A

Example Output from USEPA BMD Program

```
============================================================
        Multistage Model. $Revision: 2.1 $ $Date: 2000/08/21 03:38:21 $
        Input Data File: D:\BMDS\DATA\DICHOTOMOUS.(d)
        Gnuplot Plotting File:  D:\BMDS\DATA\DICHOTOMOUS.plt
                                                Fri Oct 08 11:40:12 2004
============================================================
BMDS MODEL RUN
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
   The form of the probability function is:

  P[response] = background + (1-background)*[1-EXP(
-beta1*dose^1-beta2*dose^2-beta3*dose^3-beta4*dose^4)]

  The parameter betas are restricted to be positive

  Dependent variable = EFFECT1
  Independent variable = DOSE

Total number of observations = 5
Total number of records with missing values = 0
Total number of parameters in model = 5
Total number of specified parameters = 0
Degree of polynomial = 4

Maximum number of iterations = 250
Relative Function Convergence has been set to: 1e-008
Parameter Convergence has been set to: 1e-008

            Default Initial Parameter Values
              Background =           0
                 Beta(1) =           0
                 Beta(2) = 1.42587e-005
                 Beta(3) = 2.17177e-007
                 Beta(4) =           0

       Asymptotic Correlation Matrix of Parameter Estimates

       ( *** The model parameter(s)  -Background    -Beta(1)    -Beta(4)
            have been estimated at a boundary point, or have been specified by the user,
            and do not appear in the correlation matrix )

           Beta(2)      Beta(3)

  Beta(2)          1       -0.97

  Beta(3)       -0.97           1
                  Parameter Estimates

     Variable         Estimate          Std. Err.
   Background                0             NA
     Beta(1)                 0             NA
     Beta(2)       1.05563e-005      2.11963e-005
     Beta(3)       2.3908e-007       1.29989e-007
     Beta(4)                 0             NA

NA - Indicates that this parameter has hit a bound implied by some inequality constraint and thus has no standard
error.
```

Analysis of Deviance Table

Model	Log(likelihood)	Deviance	Test DF	P-value
Full model	-178.191			
Fitted model	-178.237	0.0919138	3	0.9928
Reduced model	-332.032	307.682	4	<.0001

AIC: 360.474

Goodness of Fit

	Dose	Est._Prob.	Expected	Observed	Size	Chi^2 Res.
i: 1	0.0000	0.0000	0.000	0	100	0.000
i: 2	50.0000	0.0547	5.472	5	100	-0.091
i: 3	100.0000	0.2915	29.153	30	100	0.041
i: 4	150.0000	0.6481	64.810	65	100	0.008
i: 5	200.0000	0.9032	90.318	90	100	-0.036

Chi-square = 0.09 DF = 3 P-value = 0.9929

Benchmark Dose Computation

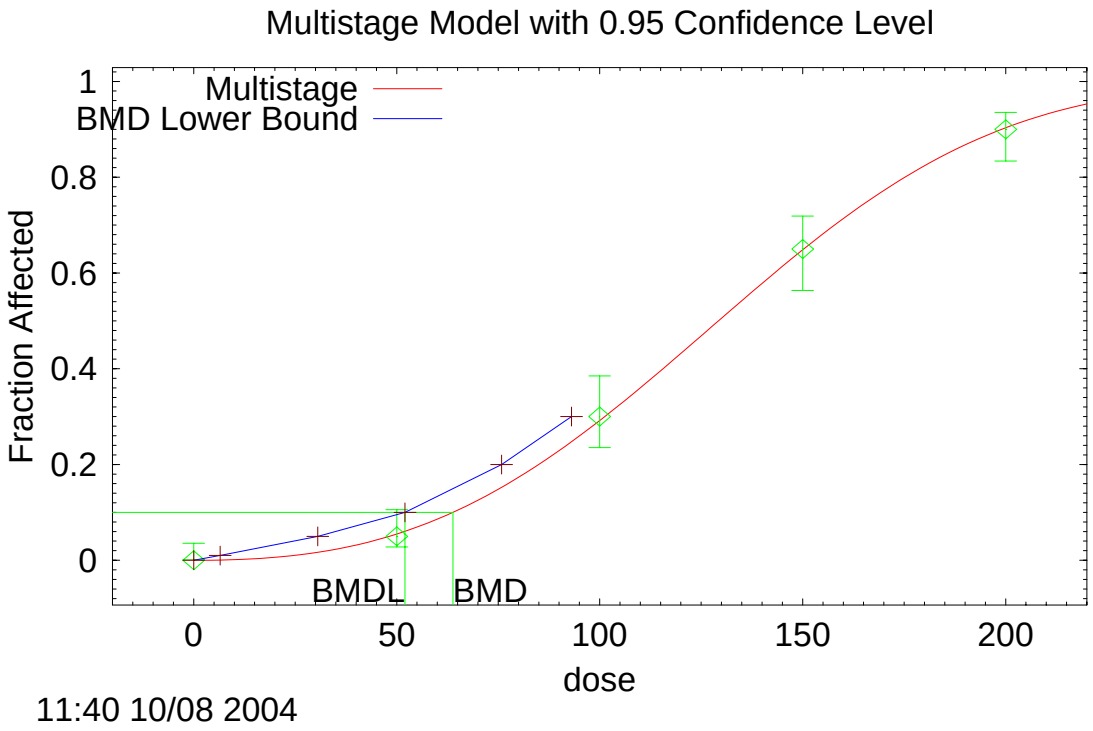
Specified effect = 0.1

Risk Type = Extra risk

Confidence level = 0.95

BMD = 63.8712

BMDL = 52.0372


Multistage Model with 0.95 Confidence Level
Multistage
BMD Lower Bound
Fraction Affected
1
0.8
0.6
0.4
0.2
0
BMDL
BMD
0
50
100
150
200
dose
11:40 10/08 2004